Definition Information $I(p)$

Definition $I(p)$

Die Information $I(p)$ eines Symbols mit Quellws $p > 0$ ist definiert als

$$I(p) = \log \frac{1}{p}.$$

Die Einheit der Information bezeichnet man als Bit.

Beispiele für Information

- $Q = \{0, 1\}$ mit $p_1 = p_2 = \frac{1}{2}$. Dann ist $I(\frac{1}{2}) = 1$, d.h. für jedes gesendete Symbol erhält der Empfänger 1 Bit an Information.
- $Q = \{0, 1\}$ mit $p_1 = 1, p_2 = 0$. Dann ist $I(1) = 0$, d.h. der Empfänger enthält 0 Bit an Information pro gesendetem Zeichen.
- Beamer-Bild SXGA: Auflösung $1280 * 1024$, 256 Farben
  - $2^{1280*1024*8}$ mögliche Bilder. Annahme: Jedes gleich wahrscheinlich.
  - Information in Bit: $I(2^{-1280*1024*8}) = 1280 * 1024 * 8 = 10.485.760$
- Meine Erklärung dieser Folie:
  $\leq$ 1000 Worte, $\leq$ 10.000 Worte Vokabular
  - Information meiner Erklärung: $I(10.000^{-1000}) < 13.288$
  - Beispiel für “Ein Bild sagt mehr als 1000 Worte!”

Entropie einer Quelle

Definition Entropie einer Quelle

Sei Q eine Quelle mit Quellws $P = \{p_1, \ldots, p_n\}$. Die *Entropie von* Q ist definiert als

$$H(Q) = \sum_{i=1}^{n} p_i I(p_i) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \log p_i.$$

- D.h. Entropie ist die erwartete Information pro Quellsymbol.
- Für $p_i = 0$ definieren wir $p_i \log \frac{1}{p_i} = 0$.
- $P = \{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\} : H(Q) = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n$
- $P = \{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}, 0\} : H(Q) = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n$
- $P = \{1, 0, 0, \ldots, 0\} : H(Q) = 1 * \log 1 = 0$

Wollen zeigen: $0 \leq H(Q) \leq \log n$.

Wechsel zu anderer Ws-Verteilung

Lemma Wechsel Ws-Verteilung

Sei $P = \{p_1, \ldots, p_n\}$ eine Ws-Verteilung und $Q = \{q_1, \ldots, q_n\}$ mit $\sum_{i=1}^{n} q_i \leq 1$. Dann gilt

$$\sum_{i=1}^{n} p_i I(p_i) \leq \sum_{i=1}^{n} p_i I(q_i).$$

Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle $i = 1, \ldots, n$.

Beweis:
Nützliche Ungleichung für das Rechnen mit logs:

$$x - 1 \geq \ln x = \log x \cdot \ln 2 \qquad \text{für alle } x > 0$$

Gleichheit gilt gdw $x = 1$.

Beweis des Lemmas

$$\begin{aligned}
\sum_{i=1}^{n} p_i I(p_i) - \sum_{i=1}^{n} p_i I(q_i) &= \sum_{i=1}^{n} p_i \left(\log \frac{1}{p_i} - \log \frac{1}{q_i} \right) \\
&= \sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \\
&\leq \frac{1}{\ln 2} \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1 \right) \\
&= \frac{1}{\ln 2} \left(\sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i \right) \\
&= \frac{1}{\ln 2} \left(\sum_{i=1}^{n} q_i - 1 \right) \leq 0.
\end{aligned}$$

Gleichheit gilt gdw $\frac{q_i}{p_i} = 1$ für alle $i = 1, \ldots, n$.

Untere und obere Schranken für $H(P)$

Satz Schranken für $H(P)$

Sei Q eine Quelle mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Dann gilt

$$0 \leq H(Q) \leq \log n.$$

Weiterhin gilt $H(Q) = \log n$ gdw alle $p_i = \frac{1}{n}$ für $i = 1, \ldots, n$ und $H(Q) = 0$ gdw $p_i = 1$ für ein $i \in [n]$.

Beweis:

- Sei $P' = \{\frac{1}{n}, \ldots, \frac{1}{n}\}$ die Gleichverteilung.
- Nach Lemma zum Wechsel von Ws-Verteilungen gilt

$$H(Q) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log \frac{1}{p'_i} = \log n \sum_{i=1}^{n} p_i = \log n.$$

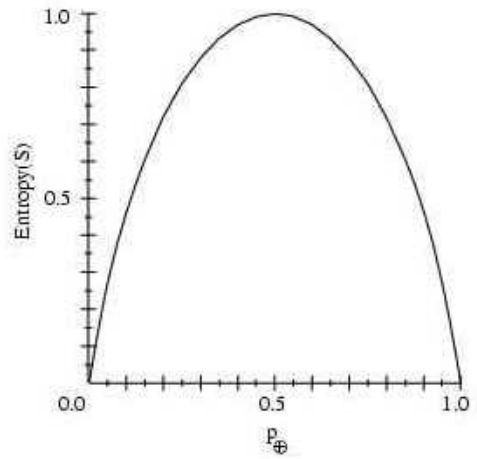
- Gleichheit gilt gdw $p_i = p'_i = \frac{1}{n}$ für alle i.

Untere Schranke für $H(P)$

Verwenden Ungleichung $\log x \geq 0$ für $x \geq 1$. Gleichheit gilt gdw $x = 1$.

$$H(Q) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \geq 0,$$

mit Gleichheit gdw $\frac{1}{p_i} = 1$ für ein $i \in [n]$. □



Bsp: binäre Entropiefunktion

- Binäre Quelle $Q = \{a_1, a_2\}$ mit $P = \{p, 1-p\}$

$$H(Q) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

Kodieren einer binären Quelle

Szenario: Binäre Quelle Q mit $P = \{\frac{1}{4}, \frac{3}{4}\}$ mit

$$H(Q) = \frac{1}{4} \cdot \log 4 + \frac{3}{4} \cdot \log \frac{4}{3} \approx 0.811.$$

- Huffman-Kodierung von Q:
$$C(a_1) = 0, C(a_2) = 1 \text{ mit } E(C) = 1.$$
- **Problem:** Wie können wir a_2 mit kurzem Codewort kodieren?
- **Idee:** Kodieren Zweierblöcke von Quellsymbolen.

Quellerweiterungen von Q

- Betrachten $Q^2 = \{a_1a_1, a_1a_2, a_2a_1, a_2a_2\}$ mit Quellws

$$p_1 = \frac{1}{16}, p_2 = p_3 = \frac{3}{16}, p_4 = \frac{9}{16}.$$

- Huffmann-Kodierung von Q^2 liefert

$$C(a_1a_1) = 101, C(a_2a_1) = 100, C(a_1a_2) = 11, C(a_2a_2) = 0$$

mit $E(C) = 3 \cdot (\frac{1}{16} + \frac{3}{16}) + 2 \cdot \frac{3}{16} + 1 \cdot \frac{9}{16} = \frac{27}{16}$.

- Jedes Codewort kodiert zwei Quellsymbole, d.h. die durchschnittliche Codewortlänge pro Quellsymbol ist

$$E(C)/2 = \frac{27}{32} = 0.844.$$

- *Übung:* Für Q^3 erhält man 0.823.

k-te Quellerweiterung Q^k

Definition k-te Quellerweiterung

Sei Q eine Quelle mit Alphabet $A = \{a_1, \ldots, a_n\}$ und Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Die *k-te Quellerweiterung* Q^k von Q ist definiert über dem Alphabet A^k, wobei $a = a_{i_1} \ldots a_{i_k} \in A^k$ die Quellws $p_{i_1} \cdot p_{i_2} \cdot \cdots \cdot p_{i_k}$ besitzt.

Satz Entropie von Q^k

Sei Q eine Quelle mit k-ter Quellerweiterung Q^k. Dann gilt

$$H(Q^k) = k \cdot H(Q).$$

Beweis für $H(Q^k)$

$$\begin{aligned} H(Q^k) &= \sum_{(i_1,\ldots,i_k)\in[n]^k} p_{i_1}\ldots p_{i_k} \log \frac{1}{p_{i_1}\ldots p_{i_k}} \\ &= \sum_{(i_1,\ldots,i_k)\in[n]^k} p_{i_1}\ldots p_{i_k} \log \frac{1}{p_{i_1}} + \cdots + \sum_{(i_1,\ldots,i_k)\in[n]^k} p_{i_1}\ldots p_{i_k} \log \frac{1}{p_{i_k}} \end{aligned}$$

- Betrachten ersten Summanden

$$\begin{aligned} \sum_{(i_1,\ldots,i_k)\in[n]^k} p_{i_1}\ldots p_{i_k} \log \frac{1}{p_{i_1}} &= \sum_{i_1\in[n]} p_{i_1} \log \frac{1}{p_{i_1}} \cdot \sum_{i_2\in[n]} p_{i_2} \cdot \ldots \cdot \sum_{i_k\in[n]} p_{i_k} \\ &= \sum_{i_1\in[n]} p_{i_1} \log \frac{1}{p_{i_1}} \cdot 1 \cdot \ldots \cdot 1 = H(Q). \end{aligned}$$

- Analog liefern die anderen $k-1$ Summanden jeweils $H(Q)$.

Kodierungstheorem von Shannon

Kodierungstheorem von Shannon (1948)

Sei Q eine Quelle für $\{a_1, \ldots, a_n\}$ mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Sei C ein kompakter Code für Q. Dann gilt für die erwartete Codewortlänge

$$H(Q) \leq E(C) < H(Q) + 1.$$

Beweis: $H(Q) \leq E(C)$

- Bezeichnen Codewortlängen $\ell_i := |C(a_i)|$ und $q_i := 2^{-\ell_i}$.
- Nach Satz von McMillan gilt: $\sum_{i=1}^{n} q_i = \sum_{i=1}^{n} 2^{-\ell_i} \leq 1$.
- Lemma Wechsel Ws-Verteilung liefert

$$\begin{aligned} H(Q) &= \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log \frac{1}{q_i} \\ &= \sum_{i=1}^{n} p_i \log 2^{\ell_i} = \sum_{i=1}^{n} p_i \ell_i = E(C). \end{aligned}$$

$E(C) \leq H(Q) + 1$

- Wir konstruieren aus $p_1, \ldots, p_n$ einen Code C'.
- Die Codewortlängen ℓ_i von C' werden wie folgt gewählt

$$\log \frac{1}{p_i} \leq \ell_i < \log \frac{1}{p_i} + 1.$$

  Ein Code C' mit dieser Eigenschaft heißt *Shannon-Fano Code*.
- Damit gilt

$$\sum_{i=1}^{n} 2^{-\ell_i} \leq \sum_{i=1}^{n} 2^{-\log \frac{1}{p_i}} = \sum_{i=1}^{n} p_i = 1.$$

- Nach dem Satz von McMillan existiert ein eindeutig entschlüsselbarer Code C' mit diesen Codewortlängen ℓ_i.
- Für jeden kompakten Code C gilt andererseits

$$\begin{aligned} E(C) \leq E(C') &= \sum_{i=1}^{n} p_i \ell_i < \sum_{i=1}^{n} p_i \left(\log \frac{1}{p_i} + 1 \right) \\ &= \sum_{i=1}^{n} p_i \log \frac{1}{p_i} + \sum_{i=1}^{n} p_i = H(Q) + 1 \end{aligned}$$

Anwendung auf Quellerweiterungen

Korollar zu Shannons Kodierungstheorem

Sei Q eine Quelle mit k-ter Quellerweiterung Q^k. Sei C ein kompakter Code für Q^k. Dann gilt

$$H(Q) \leq \frac{E(C)}{k} < H(Q) + \frac{1}{k}.$$

Beweis:

- Anwendung von Shannon's Kodierungstheorem auf Q^k liefert

$$H(Q^k) \leq E(C) < H(Q^k) + 1.$$

- Anwenden von $H(Q^k) = kH(Q)$ und teilen durch k liefert die Behauptung.

Szenario für fehlerkorrigierende Codes

Definition (n, M)-Code

Sei $C \subseteq \{0,1\}^n$ ein binärer Blockcode der Länge n mit $|C| = M$ Codeworten. Dann bezeichnen wir C als (n, M)-Code.

Erinnerung: Binärer symmetrischer Kanal

- Bits 0,1 kippen mit Ws $p, p < \frac{1}{2}$ zu 1,0.
- Korrekte Übertragung $0 \mapsto 0$, $1 \mapsto 1$ mit Ws $1 - p$.
- Kanal ist erinnerungslos, d.h. die Ws sind unabhängig von vorigen Ereignissen.
- Vorwärts-Kanalws: $\mathrm{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c} \text{ gesendet }]$.
- Rückwärts-Kanalws: $\mathrm{Ws}[\mathbf{c} \text{ gesendet } | \mathbf{x} \text{ empfangen}]$.

Dekodieren

Definition Dekodier-Kriterium

Sei $C \subseteq \{0,1\}^n$ ein (n, M)-Code. Ein Dekodier-Kriterium f ist eine Funktion $f : \{0,1\}^n \to C \cup \{\perp\}$.

- Sei $\mathbf{x} \in \{0,1\}^n$. Ein Dekodier-Kriterium dekodiert $\mathbf{x}$ zu $f(\mathbf{x}) \in C$ oder gibt Dekodierfehler $f(\mathbf{x}) = \perp$ aus.
- **Ziel:** Konstruktion eines Dekodier-Kriteriums f, dass die Ws des korrekten Dekodierens maximiert.

$$\text{Ws}[\text{korrekte Dekodierung} \mid \mathbf{x} \text{ empfangen}] = \text{Ws}[f(\mathbf{x}) \text{ gesendet} \mid \mathbf{x} \text{ empfange}$$

- Summieren über alle möglichen empfangenen $\mathbf{x}$ liefert

$$\text{Ws}[\text{korr. Dekodierung}] = \sum_{\mathbf{x} \in \{0,1\}^n} \text{Ws}[f(\mathbf{x}) \text{ gesendet} \mid \mathbf{x} \text{ empfangen}] \cdot \text{Ws}[\mathbf{x} \text{ empfangen}]$$

- Wir maximieren die Rückwärts-Kanalws

$$\text{Ws}[f(\mathbf{x}) \text{ gesendet} \mid \mathbf{x} \text{ empfangen}].$$

- D.h. wir dekodieren zu demjenigen Codewort $\mathbf{c} = f(\mathbf{x})$, das mit höchster Ws gesendet wurde.

Maximum Likelihood Dekodierung

Definition Maximum Likelihood Dekodierung

Ein Dekodierkriterium f heißt *Maximum-Likelihood Kriterium*, falls es die Vorwärts-Ws für alle Codeworte maximiert, d.h.

$$\mathrm{Ws}[\mathbf{x} \text{ empfangen } | f(\mathbf{x}) \text{ gesendet}] = \max_{c \in C} \mathrm{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c} \text{ gesendet }].$$

Eine Anwendung von f heißt Maximum-Likelihood Dekodierung.

Warum Maximum Likelihood?

Satz Maximum Likelihood optimal für gleichverteilte Codeworte

Sei C ein (n, M)-Code und $\text{Ws}[\mathbf{c} \text{ gesendet}] = \frac{1}{M}$ für alle $\mathbf{c} \in C$. Dann maximiert die Maximum-Likelihood Dekodierung die Ws des korrekten Dekodierens.

Beweis:

$$\text{Ws}[\mathbf{c} \text{ gesendet } | \mathbf{x} \text{ empfangen }]$$

$$= \frac{\text{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c} \text{ gesendet}]\text{Ws}[\mathbf{c} \text{ gesendet}]}{\sum_{i=1}^{M} \text{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c_i} \text{ gesendet}]\text{Ws}[\mathbf{c_i} \text{ gesendet}]}$$

$$= \frac{\text{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c} \text{ gesendet}]}{\sum_{i=1}^{M} \text{Ws}[\mathbf{x} \text{ empfangen } | \mathbf{c_i} \text{ gesendet}]}$$

- Nenner ist konstant für jeden Kanal.
- Maximum-Likelihood maximiert den Zähler und damit den Term.

Dekodieren zum Nachbarn minimaler Distanz

Definition Hamming-Distanz

Seien $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$. Die Hamming-Distanz $d(\mathbf{x}, \mathbf{y})$ ist definiert als die Anzahl der Stellen, an denen sich **x** und **y** unterscheiden.

Satz

In jedem binären symmetrischen Kanal ist das Dekodier-Kriterium, das ein **x** zum Codewort minimaler Hamming-Distanz dekodiert ein Maximum-Likelihood Kriterium.

Beweis:

- Ws von genau k Fehlern an festen Stellen beim Senden von **c**

$$\mathrm{Ws}[\mathbf{x} \text{ empfangen} | \mathbf{c} \text{ gesendet}] = p^k (1-p)^{n-k}.$$

- Wegen $p < \frac{1}{2}$ gilt $p < 1 - p$. Ein Dekodierkriterium f, das ein Codewort **c** mit minimaler Distanz $d(\mathbf{x}, \mathbf{c})$ wählt, minimiert k .
- Damit maximiert f die Vorwärts-Kanalws und ist somit ein Maximum-Likelihood Kriterium.